The Female Sexual Function Index (FSFI)—A Systematic Review of Measurement Properties

Neijenhuijs, Koen I.; Hooghiemstra, Nienke; Holtmaat, Karen; Aaronson, Neil K.; Groenvold, Mogens; Holzner, Bernhard; Terwee, Caroline B.; Cuijpers, Pim; Verdonck-de Leeuw, Irma M.

published in
Journal of Sexual Medicine
2019

DOI (link to publisher)
10.1016/j.jsxm.2019.03.001

document version
Publisher's PDF, also known as Version of record

THE JOURNAL OF
SEXUAL MEDICINE

REVIEW

The Female Sexual Function Index (FSFI)—A Systematic Review of Measurement Properties

Koen I. Neijenhuijs, Msc,[1,2] Nienke Hooghiemstra, Msc,[1,2] Karen Holtmaat, Msc,[1,2] Neil K. Aaronson, PhD,[4] Mogens Groenvold, MD, PhD,[5,6] Bernhard Holzner, PhD,[7] Caroline B. Terwee, PhD,[8] Pim Cuijpers, PhD,[1] and Irma M. Verdonck-de Leeuw, PhD[1,2,3]

ABSTRACT

Introduction: The Female Sexual Function Index (FSFI) is a patient-reported outcome measure measuring female sexual dysfunction. The FSFI-19 was developed with 6 theoretical subscales in 2000. In 2010, a shortened version became available (FSFI-6).

Aim: To investigate the measurement properties of the FSFI-19 and FSFI-6.

Methods: A systematic search was performed of Embase, Medline, and Web of Science for studies that investigated measurement properties of the FSFI-19 or FSFI-6 up to April 2018. Data were extracted and analyzed according to COnsensus-based Standards for the selection of health Measurement INstruments (COSMIN) guidelines. Evidence was categorized into sufficient, insufficient, inconsistent, or indeterminate, and quality of evidence as very high, high, moderate, or low.

Main Outcome Measures: The Main Outcome Measure is the evidence of a measurement property, and the quality of evidence based on the COSMIN guidelines.

Results: 83 studies were included. Concerning the FSFI-19, the evidence for internal consistency was sufficient and of moderate quality. The evidence for reliability was sufficient but of low quality. The evidence for criterion validity was sufficient and of high quality. The evidence for structural validity was inconsistent of low quality. The evidence for construct validity was inconsistent of moderate quality. Concerning the FSFI-6, the evidence for criterion validity was sufficient of moderate quality. The evidence for internal consistency was rated as indeterminate. The evidence for reliability was inconsistent of low quality. The evidence for construct validity was inconsistent of very low quality. No information was available on structural validity of the FSFI-6, and measurement error, responsiveness, and cross-cultural validity of both FSFI-6 and FSFI-19.

Clinical Implications: Conflicting and lack of evidence for some of the measurement properties of the FSFI-19 and FSFI-6 indicates the importance of further research on the validity of these patient-reported outcome measures. We advise researchers who use the FSFI-19 to perform confirmatory factor analyses and report the factor structure found in their sample. Regardless of these concerns, the FSFI-19 and FSFI-6 have strong criterion validity. Pragmatically, they are good screening tools for the current definition of female sexual dysfunction.

Strength & Limitation: A strong point of the review is the use of predefined guidelines. A limitation is the use of a precise rather than a sensitive search filter.

Conclusions: The FSFI requires more research on structural validity (FSFI-19 and FSFI-6), reliability (FSFI-6), construct validity (FSFI-19), measurement error (FSFI-19 and FSFI-6), and responsiveness

Received November 29, 2018. Accepted March 2, 2019.

[1]Vrije Universiteit Amsterdam, Department of Clinical, Neuro- and Developmental Psychology, Amsterdam Public Health Research Institute, Amsterdam, The Netherlands;

[2]Amsterdam UMC, Cancer Center Amsterdam, Amsterdam, The Netherlands;

[3]Amsterdam UMC, Department of Otolaryngology-Head and Neck Surgery, Amsterdam Public Health Research Institute, Amsterdam, The Netherlands;

[4]Division of Psychosocial Research and Epidemiology, The Netherlands Cancer Institute, Amsterdam, The Netherlands;

[5]The Research Unit, Department of Palliative Medicine, Bispebjerg Hospital, Copenhagen University Hospital, Copenhagen, Denmark;

[6]Department of Public Health, University of Copenhagen, Copenhagen, Denmark;

[7]Department of Psychiatry, Psychotherapy and Psychosomatics, CL-Service, Medical University of Innsbruck, Innsbruck, Austria;

[8]Amsterdam UMC, Vrije Universiteit Amsterdam, Department of Epidemiology and Biostatistics, Amsterdam Public Health Research Institute, Amsterdam, The Netherlands

https://doi.org/10.1016/j.jsxm.2019.03.001

(FSFI-19 and FSFI-6). Further corroboration of measurement invariance (both across cultures and across subpopulations) in the factor structure of the FSFI-19 is necessary, as well as tests for the unidimensionality of the FSFI-6. **Neijenhuijs KI, Hooghiemstra N, Holtmaat K, et al. The Female Sexual Function Index (FSFI)—A Systematic Review of Measurement Properties. J Sex Med 2019;16:640−660.**

Key Words: Female Sexual Function Index; Validity; Reliability; Female Sexual Dysfunction; Patient Reported Outcome Measure

INTRODUCTION

Sexual dysfunction refers to a problem that prevents people experiencing satisfaction from sexual activity. A first model of female sexual dysfunction (FSD) was composed in 1998 with 4 categories of disorders (desire, arousal, orgasm, and pain) as described in the Diagnostic and Statistical Manual of Mental Disorders-IV (DSM-IV) and the International Statistical Classification of Diseases and Related Health Problems-10 (ICD-10) at that time.[1]

In 2000, the Female Sexual Function Index (FSFI) was developed to measure female sexual (dys)function,[2] based on the models described in the DSM-IV and ICD-10. The FSFI is a 19-item patient-reported outcome measure (PROM), consisting of 6 separate domains of female sexual function, namely desire (items 1−2), arousal (3−6), lubrication (7−10), orgasm (11−13), satisfaction (14−16), and pain (17−19). Initial validation showed good internal consistency for all scales in a study sample drawn from the general population, as well as in subgroups of FSD patients and control subjects (Cronbach's α = 0.82−0.97). Test-retest reliability was acceptable (r = 0.79−0.86). Known-groups comparison was tested between FSD patients and control subjects, with significant differences on all domains of the FSFI-19. Divergent validity (as measured with the Locke-Wallace Marital Adjustment Test[3]) was good (r = 0.04−0.43), except for the FSFI satisfaction scale (r = 0.40−0.72).[2] In 2010, a 6-item version (FSFI-6) to measure FSD was developed. The 6 items were selected by inspecting the receiver operating characteristic curves of each item of the FSFI-19 for distinguishing between women with and without FSD. The best-performing item for each of the 6 domains of the FSFI-19 was selected for use in the FSFI-6.[4] The FSFI-6 showed acceptable internal consistency (Cronbach's α = 0.789), acceptable test-retest reliability (Pearson correlation = 0.95), and good criterion validity with a cutoff of $\leq$ 19 (sensitivity = 0.96; specificity = 0.91).

With the release of the DSM-5 in 2013, the model for FSD has seen some changes. Of particular interest, 1 desire disorder (sexual aversion disorder) was removed, whereas the remaining desire disorder (hypoactive desire disorder) was merged with the arousal dysfunction disorder.[5] This new model suggests that desire and arousal may not be separate constructs in the context of FSD. Interestingly, the original validation study found a 5-factor structure where desire and arousal were part of the same construct. This factor was split into 2 subscales due to clinical considerations.[2]

The FSFI-19 and FSFI-6 are widely used in clinical practice as a screening tool for FSD, as well as in clinical trials as an outcome measure. As such, it is of importance to assess the measurement properties of the FSFI-19 and FSFI-6, to determine whether they are fit to use in clinical and scientific contexts. To our knowledge, the measurement properties of the FSFI-19 and FSFI-6 have not yet been systematically reviewed. As such, the aim of this study was to investigate whether the initial good results regarding the measurement properties of the FSFI-19 and FSFI-6 were confirmed in later studies. Of particular interest is structural validity, and the question is whether the original 6-factor structure is challenged in favor of a 5-factor structure where desire and arousal are part of the same construct. The results of this systematic review are relevant for the use of the FSFI to monitor sexual dysfunction in women in clinical trials and practice.

In this review, we followed the COnsensus-based Standards for the selection of health Measurement INstruments (COSMIN) methodology.[6] This methodology is based on a taxonomy and definitions of measurement properties for PROMs[7] including content validity (assessment of whether the FSFI represents all facets of FSD), structural validity (assessment of whether the FSFI subscales are singular constructs), internal consistency (assessment of whether FSFI items measuring the same construct are consistent in their results), cross-cultural validity (assessment of whether there are structural differences in validity of the FSFI between populations), reliability (assessment of whether the FSFI reproduces similar scores when FSD has been stable), measurement error (assessment of systematic and random error between the FSFI score and the true score of a patient), criterion validity (assessment of how well the FSFI score is an adequate reflection of FSD diagnosis), hypotheses testing for construct validity (assessment of whether the FSFI measures the construct of FSD), and responsiveness (assessment of how well the FSFI measures FSD change over time).

METHODS

Literature Search

The literature search was part of a larger systematic review (Prospero ID 42017057237), which investigated the measurement properties of 39 different PROMs (including the FSFI)

measuring quality of life of cancer survivors included in an eHealth application called "Oncokompas".[8–11] The databases Embase, Medline, and Web of Science were searched using the search terms of the PROM's name and acronyms, combined with a precise search filter for measurement properties.[12] The search was performed in January 2017. A search update was performed on April 13, 2018, to search for recent studies. This search update also used broader search terms across all years (not only 2017–2018), because not all acronyms of the FSFI were correctly specified in the original search. Appendix A contains the full search terms. Due to the limitation of the sensitivity of the precise filter (93% sensitive),[12] a manual search using rudimentary search filters was performed in Google Scholar and PubMed to check for any prominent records missed in the search update.

Inclusion and Exclusion Criteria

Studies were included when they reported original data on ≥1 of the following measurement properties of the FSFI: structural validity (whether the hypothesized measurement model is confirmed), internal consistency (the degree of interrelatedness among the items of the measure), reliability (the proportion of total variance between multiple measurements, which is due to "true" differences between measurements), measurement error (a measure of systematic and random error in change scores), criterion validity (whether the measure is an adequate reflection of a gold standard; in the case of the FSFI, this is most often a diagnosis of FSD), cross-cultural validity (whether the test can be interpreted similarly in different cultures), responsiveness (whether the measure is capable of measuring change over time in the construct to be measured), and hypothesis testing for construct validity (whether the test measures the construct it proposes to measure), which consists of known-groups comparison (a comparison between groups known to have differences on the construct), convergent validity (correlations with other measures that should be related), and divergent validity (correlations with other measures that should be unrelated). While of importance for establishing validity, content validity was not investigated because it was beyond the scope of the current review. Validation studies on other PROMs that also reported original data on the FSFI were included as well.

Studies that were only available as abstracts or conference proceedings were excluded, as well as non-English publications. Titles and abstracts, and the selected full-texts were screened by 2 independent reviewers (K.N./M.V./K.H./N.H.). Disagreements were discussed until consensus was reached.

Data Extraction

Data on each of the measurement properties defined by the COSMIN taxonomy[7] was extracted by 2 independent researchers (K.N./A.vdH./H.M./E.V./N.H.). Relevant data included the type of measurement property, its results, and information on missing values. Information on the type of research (psychometric or not), specified research aim, sample size, population information, and which version of the FSFI was used, was also extracted. Disagreements were discussed until consensus was reached.

Data Analysis

Data analysis consisted of 3 consecutive steps. First, the quality of the included studies was rated using the 4-point scoring system of the COSMIN checklist.[13] Methodologic aspects regarding design requirements and preferred statistical methods, specific to each measurement property under consideration, were rated as either "inadequate," "doubtful," "adequate," or "very good." The methodologic quality was summarized per measurement property per study, as the lowest score received on any of the methodologic aspects. The complete criteria for study quality per measurement property are documented elsewhere.[13] Appendix B contains the final study quality ratings.

Second, each measurement property in each individual study was rated as sufficient, insufficient, or indeterminate, according to criteria for good measurement properties included in the COSMIN guidelines for systematic reviews of PROMs. The complete criteria for rating these measurement properties are documented elsewhere.[6] These ratings were qualitatively summarized to determine the overall rating of the measurement property for the FSFI. If all studies indicated a "sufficient," "insufficient," or "indeterminate" rating for a specific measurement property, the overall rating of this measurement property was rated accordingly. If there were inconsistencies between studies, explanations were explored (eg, differences in methodologic quality, differences in population, etc). If explanations were found, they were discussed until consensus was reached regarding the overall rating of the measurement property. If no explanations were found, the overall rating would be inconsistent.

Third, the overall rating of evidence per measurement property was supplemented by a level of quality of evidence, using a modified Grading of Recommendations Assessment, Development and Evaluation approach from the COSMIN methodology.[6] This approach takes into account (i) study quality, (ii) directness of evidence, (iii) inconsistency of results, and (iv) precision of evidence (number of studies and sample size). The overall quality of evidence was rated as high, moderate, low, or very low. Measurement properties that were rated as indeterminate in the previous step, did not receive a rating in this third step as there was no evidence to rate.

All ratings (methodologic quality, measurement property rating, and Grading of Recommendations Assessment, Development, and Evaluation rating) were made by 2 independent researchers (K.N./K.H./N.H.). Discrepancies in ratings were discussed until consensus was reached.

RESULTS

Search Results

The initial search identified 1,401 non-duplicate abstracts, of which 174 were relevant to the FSFI (Figure 1). A total of 155

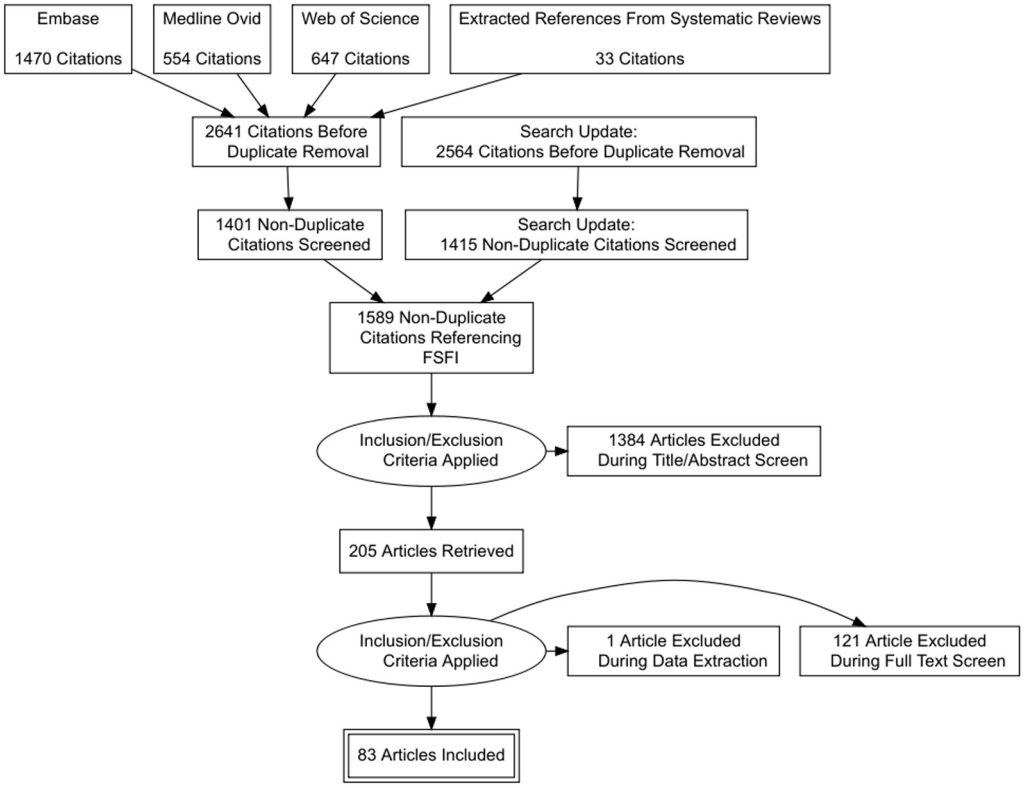

Figure 1. PRISMA diagram.

abstracts and 11 full texts were excluded from the initial search, because they did not provide unique information on a measurement property. The search update up to April 2018 identified 1,415 more non-duplicate abstracts. A total of 1,229 abstracts and 110 full-texts were excluded from the search update, because they did not provide unique information on a measurement property of the FSFI. 2 full texts were excluded during data extraction.

In total we included 83 studies: 75 on the FSFI-19,[2,14–87] 5 on the FSFI-6,[4,88–91] and 3 on adaptations of the FSFI-19: a version specific for breast cancer survivors,[92] a version for life-long sexual dysfunction,[93] and a version with an added item concerning vaginismus.[94] An overview of study characteristics is provided in Table 1.

The combined body of the 75 studies on the FSFI-19, and the 5 studies on the FSFI-6 reported on all measurement properties, except measurement error, responsiveness, and cross-cultural validity. The 3 studies on adaptations of the FSFI-19 reported on structural validity, internal consistency, test-retest reliability, and construct validity of original subscales of the FSFI-19.

Structural Validity

29 studies[2,16,19,20,24–27,33,38,39,41,44,45,47,50,58,59,63,65,73–75,78,82–85,92,93] reported on structural validity of the FSFI-19, of which 8 reported multiple analyses[2,33,50,75,78,83,85,93] (Table 2). Methodologic quality of these studies was rated as "very good"[2,19,44,47], "adequate,"[16,20,24,25,27,33,38,39,41,45,50,58,73–75,78,82–85,92,93] or "inadequate".[2,59,65] The "inadequate" ratings were due to sample sizes that were too small ("other flaws" in COSMIN methodologic quality).

3 studies of "very good" quality,[19,44,93] 1 study of "adequate" quality,[78] and 1 study of "inadequate" quality[59] confirmed the hypothesized 6-factor structure and were thus rated as sufficient. 8 studies of "adequate" quality[25,33,47,50,74,78,83,85] showed a poor fit for the 6-factor structure. 2 of these studies tested and showed support for a 5-factor structure.[33,50]

19 studies of "adequate" quality,[16,20,24,27,33,38,39,41,45,50,58,63,73,75,83–85,92,93] and 2 studies of "inadequate" quality[2,65] performed analyses (mostly Principal Component Analysis) without reporting fit statistics; therefore, results were rated as indeterminate. Noteworthy is that 5 studies showed support for a 6-factor structure,[16,65,75,82,85] whereas 11 studies showed support for a 5-factor structure with a merging of the desire and arousal subscales.[2,20,24,33,38,50,58,63,82,92,93] 1 study showed support for a 5-factor structure with a different merging of subscales,[39] and 7 studies showed support for <5 factors.[27,41,45,73,82–84] 1 study used item response theory analysis and was rated indeterminate because no fit measures were reported.[26] None of the studies that investigated the FSFI-6 [4,88–91] reported on structural validity.

Internal Consistency

36 studies reported on internal consistency of the FSFI-19[2,14,16,19,20,27,29,33,36,38–40,44,47,48,50,52,58,59,63,65,68,70,72–76,78,80,82,83,85,86,93,94] (Supplementary Table 1). Methodologic

quality was rated as "very good,"[2,14,16,19,20,33,36,38,40,44,47,52,58,63,68,70,73–76,78,80,82,83,85,93,94] "adequate",[48,59,72] "doubtful",[29,39] or "inadequate".[27,86] The "inadequate" ratings were due to reporting of Cronbach's α values for only the total score of the FSFI-19. The "doubtful" ratings were due to the reporting of Cronbach's α for an adapted version of a subscale.

23 studies[2,14,16,19,20,33,38,40,44,47,52,58,68,70,73,75,76,78,80,82,83,93,94] of "very good" quality, 2 studies of "adequate" quality,[59,72] 1 study of "doubtful" quality,[39] and 1 study of "inadequate" quality[86] reported Cronbach's α values that were rated as sufficient ($\alpha \geq 0.70$ and ≤ 0.95) for all subscales. 4 studies of "very good" quality,[36,63,74,85] 1 study of "adequate" quality,[48] 1 study of "doubtful" quality,[29] and 1 study of "inadequate" quality[27] reported multiple Cronbach's α values that were rated as insufficient ($\alpha < 0.70$ or > 0.95).

4 studies reported on internal consistency of the FSFI-6[4,88,89,91] (Supplementary Table 1). Methodologic quality was rated as "very good"[4,88,91] or "inadequate".[89] The inadequate rating was due to unclear reporting on which items the Cronbach's α was calculated. The evidence of internal consistency was rated as indeterminate for all 4 studies, because unidimensionality of the FSFI-6 was not investigated (see Structural Validity), which is a prerequisite for interpreting internal consistency.

Test-Retest Reliability

21 studies[2,16,19,22,27,33,36,39,50,58,63,65,68,70,73–75,80,84–86] reported on test-retest reliability of the FSFI-19 (Table 3). Methodologic quality was rated as "adequate",[22,33,58,84,85] "doubtful",[2,16,19,27,36,39,50,63,65,68,70,73,80,86] or "inadequate".[74,75] The "doubtful" ratings were due to the use of Pearson correlation instead of the intraclass correlation coefficients. The "inadequate" ratings were due to a very small sample size ("other flaws" in COSMIN methodologic quality),[74] or due to dissimilar test conditions.[75]

5 studies of "adequate" quality,[22,33,58,84,85] 12 studies of "doubtful" quality,[2,16,19,36,39,50,63,65,70,73,80,86] and 2 studies of "inadequate" quality[74,75] reported test-retest values that were rated as sufficient. 2 studies of "doubtful" quality[27,68] reported test-retest values that were rated as insufficient.

2 studies reported on test-retest reliability of the FSFI-6[4,89] (Table 3). Methodologic quality was rated as "adequate",[89] or "doubtful".[4] The "doubtful" rating was due to use of Pearson correlation instead of the intraclass correlation coefficient. 1 study of "doubtful" quality[4] reported test-retest values that were rated as sufficient. 1 study of "adequate" quality reported test-retest values that were rated as insufficient.

Construct Validity (Hypothesis Testing)

Known-group comparison

23 studies[2,14,16,20,29,33,39,48,52,53,58,64,65,68,70,73–76,80,82,85,86] reported on known-group comparison of the FSFI-19 (Supplementary Table 2). Known-group differences were investigated in relation to urologic/gynecologic patients vs controls,[14,39,48,80] FSD patients vs controls,[2,16,33,53,58,70,73,75,82] cancer treatment modality,[20] patients with hypoactive sexual desire disorder (HSDD) vs controls,[29,52,64,68] diabetic patients vs controls,[65] premenopausal women vs postmenopausal women,[74,85] age,[76] marriage status,[85] and women experiencing subjective sexual distress vs controls.[86] Methodologic quality was rated as "adequate" for all 23 studies. In all 23 studies the known-group comparisons provided evidence of sufficient construct validity, as the hypothesized differences between groups were confirmed. None of the studies that investigated the FSFI-6[4,88–91] reported on known-group comparisons.

Convergent Validity

49 studies reported on convergent validity of the FSFI-19 (Supplementary Table 3). The FSFI-19 was compared to measures measuring sexual function and satisfaction,[15,17,18,21,23–25,28,29,31,33,37,46,53,55,57,58,60,62,64,67–69,71,72,74,78,79,92,93] quality of life,[19,92] mental health measures,[20,39] physical functional problems,[20,30,32,34,49,54,66,74,77,81,83,87] relationship quality,[20,39,52] and body image.[35,42,43,56,60] A full list of comparison measures can be found in Supplementary Table 3.

Methodologic quality was rated as "adequate,"[15,17–21,23–25,28–32,34,35,37,39,42,43,46,49,52,54–57,60,62,66,68,69,71,72,74,77–79,81,83,87,93] "doubtful",[33,58,64,92] or "inadequate".[53] The "inadequate" rating was due to serious concerns regarding the measurement properties of the comparator instrument. The "doubtful" ratings were due to concerns regarding the measurement properties of the comparator instrument. 28 studies of "adequate" rating,[15,19,23,24,28–31,35,37,46,49,54–57,60,66–69,71,72,74,77,78,81,87,93] 2 studies of "doubtful" rating,[33,64] and 1 study of "inadequate" quality,[53] provided correlations rated as sufficient. 14 studies of "adequate" rating,[17,20,21,25,32,34,39,42,43,52,60,62,79,83] and 2 studies of "doubtful" rating,[58,92] provided correlations rated as insufficient. 1 study was rated as indeterminate, because not enough information was given for a reliable interpretation.[18]

4 studies reported on convergent validity of the FSFI-6 (Supplementary Table 3). The FSFI-6 was compared to coital frequency,[88] educational level,[88] partner educational level,[88] age,[88] partner age,[88] waist circumference,[88] hot flush intensity,[88] FSFI-19,[89] British National Survey of Sexual Attitudes and Lifestyles—Sexual Function,[90] Menopause Rating Scale,[91] and Hospital Anxiety and Depression Scale.[91] Methodologic quality was rated as "adequate",[89–91] or "doubtful".[88] The "doubtful" rating was due to concerns regarding the measurement properties of the comparator instrument. 2 studies of "adequate" quality[89,90] reported correlations rated as sufficient. 1 study of "adequate" quality,[91] and 1 study of "doubtful" quality[88] reported correlations rated as insufficient.

Divergent Validity

8 studies reported on divergent validity of the FSFI-19[2,14,19,48,58,68,76,92] (Supplementary Table 4).

Table 1. Characteristics of included studies

Reference	Population	Sample size	Main aim of study
FSFI-19			
Achimas-Cadariu et al[14]	Female patients of reproductive age treated for premalignant and malignant pathology of the uterine cervix	102/204	To investigate the associations among quality of life determinants on a sample of female patients of reproductive age (102 patients and 102 healthy control subjects), surgically treated (conization) for preinvasive and invasive pathology (cervix dysplasia and microinvasive carcinoma)
Ahmed et al[15]	Premenopausal women	241	To assess sexually related personal distress among premenopausal women with sexual dysfunction via a validated Arabic version of the original FSDS
Anis et al[16]	Egyptian women	855	To validate the ArFSFI
Aydin et al[17]	Turkish women visiting the urogynecology clinic	248	To develop a Turkish version of the FSDS-R, to evaluate its psychometric reliability and validity, and to estimate the optimal cutoff score that corresponds best to the clinical diagnosis of sexual dysfunction
Azimi Nekoo et al[18]	Married and potentially sexually active Iranian women	1,966	To determine the psychometric properties of the Iranian version of the FSDS-R in a population sample of Iranian women
Bartula et al[19]	Breast cancer patients	399	Assess extent to which FSFI is applicable for breast cancer patients
Baser et al[20]	Female cancer survivors	181	Systematic evaluation of the factor structure, reliability, and construct validity of the FSFI for measuring the sexual functioning of female cancer survivors
Bloemendaal et al[21]	Dutch women	323	Validate the Dutch version of the SESII-W
Borello-France et al[22]	Female patients with a relapsing form of MS	48	To assess the test-retest reliability of the U-UDI and the FSFI in women with MS
Bornefeld-Ettmann et al[23]	German-speaking women	465	The German translation of the SSEI-SF by Zeanah and Schwarz (1996) was validated via an online survey with 557 women and then investigated in a clinical sample of women suffering from PTSD after sexual and physical abuse compared with healthy controls
Burri et al[24]	Sexually active Swiss women	309	To evaluate the validity and utility of the German version of the SCS-W by assessing content, convergent, and discriminant validity
Carpenter et al[25]	Midlife postmenopausal women	93	To evaluate whether a single item from the FSDS-R could be identified to use to screen midlife women for bothersome diminution in sexual function
Carpenter et al[26]	Peri- and postmenopausal women reporting hot flashes	898	To evaluate whether a subset of items on the 19-item English-language FSFI would perform as well as the full length FSFI in peri- and postmenopausal women
Chang et al[27]	Pregnant women receiving prenatal examinations	108	To translate the FSFI from English to traditional Chinese, and to evaluate the reliability and validity of this new version for pregnant women
Clayton et al[28]	Female patients with diagnosis of HSDD	90	To assess the reliability and validity of the SIDI-F as a measure of HSDD severity
Clayton et al[29]	American (N = 220) and European (N = 253) women going through FSD diagnosis	473	To estimate the reliability and validity of the SIDI-F as a measure of HSDD severity
Constantine et al[30]	American and British women	589	To create a valid and responsive summary score for the PISQ-IR
DeRogatis et al[31]	Postmenopausal female patients aged 40–65 with spontaneous amenorrhea or bilateral oophorectomy with or without hysterectomy at least 6 months before study	629	To validate the WSID-SF and DSLA in postmenopausal women

(continued)

Table 1. Continued

Reference	Population	Sample size	Main aim of study
Eaton et al[32]	Female cancer survivors	175	To develop and validate brief clinical measurements to facilitate the identification of vulvovaginal symptoms in patients with and survivors of cancer
Fakhri et al[33]	Iranian gynecological outpatients	448	To translate, validate, and enhance cross-cultural comparability of the FSFI-IV
Farkas et al[34]	Female patients diagnosed with Pelvic Organ Prolapse, Urinary Incontinence, or Fecal Incontinence	178	To translate the PISQ-IR into Hungarian and to validate the translated PISQ-IR
Ferguson et al[35]	Women visiting gynecologic oncology outpatient clinic	268	To confirm the factor structure of the Sexual Adjustment and Body Image Scale using a confirmatory factor analysis
Filocamo et al[36]	Italian women visiting urological and gynecological clinics	409	To perform a linguistic validation of the Italian version of the FSFI
Flynn et al[37]	American female cancer patients	430	Validation of PROMIS sexual function and satisfaction scales
Forbes et al[38]	Sexually active Australian women	336	To examine the measurement capabilities of the IIEF and FSFI based on data collected from an online study in 2010
Ghassamia et al[39]	Iranian women	562	To examine the psychometric properties of a P-FSFI among a sample of healthy Iranian women
Heng et al[41]	Malaysian women visiting infertility clinic	150	To determine the construct of the phases of the female SRC among women attending an infertility clinic in a Malaysian tertiary center
Herbenick et al[42]	American women attending sex toy parties	1,937	To establish a reliable and valid measure of female genital self-image, the FGSIS, and to assess the relationship between scores on the FGSIS and women's sexual function
Herbenick et al[43]	American women	2,056	To assess the reliability and validity of the FGSIS, its model of fit, and its association with women's scores on the FSFI in a nationally representative probability sample of women in the United States ages 18–60
Hevesi et al[44]	202 university students, 177 patients with endometriosis, and 129 patients with polycystic ovary syndrome; from Hungary	508	To investigate whether female sexual function is best understood as a multidimensional construct or, alternatively, whether a common underlying factor explains most of the variance in FSFI scores
Ismail et al[45]	178 female patients with diabetes, and 175 women without diabetes from Malaysia	353	To compare the components of sexual responses between Malaysian women with type 2 diabetes mellitus and those without the disease
Jing et al[46]	Breast cancer survivors	246	To develop a Quality of Sexual Life Questionnaire in Breast Cancer Survivors and determine its validity and reliability
Kalmbach et al[47]	Female undergraduate students	409	To assess factor structures of the FSFI, MSFI (adapted for this investigation), and PFSF in young, healthy men and women
Likes et al[48]	43 female patients with vulvar excisions for vulvar intraepithelial neoplasia; 43 age-matched controls	86	To extend the validation of the FSFI to include women with vulvar excisions for vulvar intraepithelial neoplasia
Liu et al[49]	Chinese female patients with interstitial cystitis and bladder pain syndrome	90	To examine whether adding a sexual dysfunction domain to the UPOINT system improves the association with IC-BPS symptom severity due to a high prevalence of sexual dysfunction in women
Liu et al[50]	Female inpatients with cervical cancer	215	To examine the psychometric properties and performance of the CVFSFI among a sample of Chinese women with cervical cancer

(continued)

Table 1. Continued

Reference	Population	Sample size	Main aim of study
Ma et al[51]	Chinese women	500	To establish clinical cutoff scores for the CVFSFI and to evaluate the prevalence of FSD in urban Chinese women
Meston et al[52]	71 female patients with female orgasmic disorder, 44 female patients with hypoactive sexual desire disorder, and 71 healthy women	186	To extend the validation of the FSFI to include women with a primary clinical diagnosis of female orgasmic disorder or hypoactive sexual desire
Meston et al[53]	American women	172	To develop a comprehensive, multifaceted, valid, and reliable self-report measure of women's sexual satisfaction and distress
Mestre et al[54]	118 not sexually active women, and 150 sexually active women	268	To transculturally adapt the PISQ-IR into Spanish
Mohammadi et al[55]	Iranian married women with MS	226	To translate and validate the MSISQ-19 in women with MS in Iran
Mohammed et al[56]	Egyptian married women	244	To translate the original English version of the FGSIS into Arabic and validate the Arabic version
Nimbi et al[57]	Italian women	626	To test the psychometric characteristics of the Italian version of the SMQ focusing on the Automatic Thoughts subscale
Nowosielski et al[58]	85 Polish female patients with FSD, 104 Polish women without FSD	189	To develop a Polish version of the FSFI
Opperman et al[59]	Canadian women	85	To evaluate and compare 4 models of the FSFI: (i) single-factor model, (ii) 6-factor model, (iii) second-order factor model, and (iv) 5-factor model combining the desire and arousal subscales
Pakpour et al[60]	Iranian female population sample (n = 2,675), Iranian female patients with FSD (n = 295), Iranian female patients with type 2 diabetes (n = 449)	3,419	The purpose of this study was the translation and validation of the SQOL-F in Iranian women
Pakpour et al[61]	Iranian female students	1,877	To investigate the psychometric properties of a translated and culturally adapted FGSIS-I in a sample of college women
Pascoal et al[62]	Heterosexual sexually active women involved in a dyadic relationship	278	To describe the development and validation of the Beliefs About Sexual Functioning Scale
Rehman et al[63]	Bilingually educated women in a stable sexual relationship	116	To translate, cross-culturally adapt, and perform a psychometric validation of an Urdu translation of the FSFI
Rellini et al[64]	Female patients with female sexual arousal disorder	24	To provide empiric evidence on the sensitivity of different types of measures for detecting treatment-induced changes in FSD diagnosis
Rillon-Tabil et al[65]	Ambulatory women	85	To translate and validate the FSFI—Filipino version
Rogers et al[66]	American and British female patients with pelvic floor disorders	589	To create a valid, reliable, and responsive sexual function measure in women with PFDs for both SA and NSA women
Rosen et al[2]	Healthy women	259	
Rosen et al[67]	American women reporting normal sexual function	329	To develop and psychometrically validate a self-administered Female Sexual Well-Being Scale for assessing sexual well-being in sexually functional women

(continued)

Table 1. Continued

Reference	Population	Sample size	Main aim of study
Ryding et al[68]	50 Swedish female patients with hypoactive sexual desire disorder, and 58 age-matched healthy Swedish women	108	To investigate the psychometric properties of the Swedish version of the FSFI
Selcuk et al[69]	71 Turkish female patients with pelvic problems, and 38 Turkish healthy women	109	To validate the Turkish versions of the SHOW-Q for Turkish-speaking women
Sidi et al[70]	Married Malaysian women	230	To validate the Malay version of the FSFI
Sills et al[71]	Premenopausal female patients diagnosed with HSDD	448	To use the outcome of item response analyses of blinded data from 2 randomized, placebo-controlled trials, to assist in the revision of the scale
Stephenson et al[72]	Adult American women in a monogamous heterosexual relationship reporting sexual difficulties	97	To assess the correlations between FSFI scores and information regarding specific rates of functional impairment gained via clinical interview; and to assess the specificity of FSFI subscale scores in reflecting corresponding aspects of sexual function
Sun et al[73]	85 Chinese women seeking regular health check-up, 145 Chinese women who accompanied patients, and 98 Chinese female patients with medical illness not affecting sexual function	328	To develop and validate the CVFSFI to assess FSD in China
Takahashi et al[74]	Healthy Japanese women in partnered relationships	126	To develop the FSFI-J 3-month version, and to measure its psychometric reliability and validity
Ter Kuile et al[75]	234 Dutch female patients with FSD, and 108 Dutch women without FSD	342	To investigate the psychometric properties of the FSFI and the Female Sexual Distress Scale within a Dutch population
Trudel et al[76]	Canadian women >65 years old in a relationship	143	To validate the FSFI in an older (≥65 years), non-clinical population of francophone women living with their spouses in Quebec
Trutnovsky et al[77]	German female patients visiting urogynecologic clinics for pelvic floor dysfunction	197	To translate the PISQ-IR into German and to clinically validate it in a German-speaking population
Vallejo-Medina et al[78]	Colombian adult women	925	To adapt and validate the FSFI to Spanish language in a Colombian sample
Velten et al[79]	German adult women	2,206	To assess the psychometric properties of a German version of the SESII-W
Verit et al[80]	100 female patients with CPP and 100 age-matched women without CPP	200	To investigate the validity and reliability of FSFI in women with chronic pelvic pain
Wang et al[81]	Chinese women visiting a urogynecologic clinic	106	To translate and validate the Mandarin Chinese version of PISQ-IR for global use
Wiegel et al[82]	307 female patients with FSD diagnoses, and 261 healthy women	568	To cross-validate the FSFI in several samples of women with mixed sexual dysfunctions (n = 568) and to develop diagnostic cutoff scores for potential classification of women's sexual dysfunction
Witting et al[83]	Finnish female adult twins	2,081	To validate the FSFI in Finnish
Wolpe et al[84]	Brazilian female physical therapy students	246	To assess the psychometric properties of the FSFI applied to the VAS
Wylomanski et al[85]	French women attending gynecology consultation	512	To validate a French version of the FSFI in a sample of French women

(continued)

Table 1. Continued

Reference	Population	Sample size	Main aim of study
Zachariou et al[86]	18 Greek female patients with FSD, and 99 Greek women without FSD	117	To linguistically validate the Greek version of the FSFI
Zohre et al[87]	100 Iranian healthy women, 200 Iranian female patients suffering from urinary incontinence with or without pelvic organ prolapse	200	To translate the PISQ-12 and provide evidence for psychometric properties
FSFI-19 Sexual Desire subscale only			
Gerstenberger et al[40]	American and Canadian women	618 & 892	To define and validate a specific cut point on the SD domain for differentiating women with and without hypoactive sexual desire disorder
FSFI-19; FSFI-LL			
Burri et al[93]	British female twins	FSFI-19: 1,056; FSFI-LL: 744	To develop a modified version of the widely used FSFI-LL, and to evaluate the psychometric properties and aptness of this new version
FSFI-20 (added item for vaginismus)			
Carvalho et al[94]	Portuguese women	1,425	To test, using structural equation modeling, 5 conceptual, alternative models of female sexual function, using a sample of women with sexual difficulties and a sample of women without sexual problems
FSFI-BC (34 items)			
Bartula et al[92]	Australian breast cancer survivors	596	To determine the reliability, validity, and acceptability of the FSFI-BC
FSFI-6			
Chedraui et al[88]	Ecuadorian women	904	To assess sexual function and related factors in mid-aged Ecuadorian women
Isidori et al[4]	Women attending a screening visit for sexual and reproductive dysfunction	160	Development of short-form version of FSFI
Lee et al[89]	Korean female patients who visited outpatient center for uterine cancer	220	To evaluate the validity and reliability of the FSFI-6K
Mitchell et al[90]	1,262 population sample, and 100 patients with sexual problems. Count of women unspecified.	1362	Development of a new measure of sexual function for the third British National Survey of Sexual Attitudes and Lifestyles
Pérez-López et al[91]	Female patients attending gynecologic and obstetric healthcare facilities	179	To assess sexual function and related factors in middle-aged Spanish women

ArFSFI = Arabic version of the Female Sexual Function Index; CPP = chronic pelvic pain; CVFSFI = Chinese version of the Female Sexual Function Index; FGSIS = Female Genital Self-Image Scale; FGSIS-I = Female Genital Self-Image Scale—Iranian version; FSD = female sexual disorder; FSDS = Female Sexual Distress Scale; FSDS-R = Female Sexual Distress Scale—Revised; FSFI = Female Sexual Function Index; FSFI-6K = Female Sexual Function Index-6—Korean version; FSFI-BC = breast cancer—specific adaptation of the Female Sexual Function Index; FSFI-IV = Female Sexual Function Index— Iranian version; FSFI-J = Female Sexual Function Index—Japanese translation; FSFI-LL = Female Sexual Function Index that allows assessment of women's lifelong sexual function; HSDD = hypoactive sexual desire disorder; IC-BPS = interstitial cystitis and bladder pain syndrome; MS = multiple sclerosis; MSFI = Male Sexual Function Index; NSA = not sexually active; PFD = pelvic floor disorder; PFSF = Profile of Female Sexual Function; P-FSFI = Persian language version of the Female Sexual Function Index; PISQ-IR = Pelvic Organ Prolapse/Incontinence Sexual Questionnaire—International Urogynecology Association Revised; PISQ-12 = Pelvic Organ Prolapse/Urinary Incontinence Sexual Questionnaire; PISQ-IR = Pelvic Organ Prolapse/Incontinence Sexual Questionnaire IUGA-Revised; PTSD = post-traumatic stress disorder; SA = sexually active; SCS-W = Sexual Complaints Screener for Women; SESII-W = Sexual Excitation/Sexual Inhibition Inventory for Women; SHOW-Q = Sexual Health Outcomes in Women Questionnaire; SMQ = Sexual Modes Questionnaire; SQOL-F = Iranian version of the Sexual Quality of Life questionnaire—Female; SRC = sexual response cycle; SSEI-SF = Sexual Self-Esteem Inventory—Short Form; UPOINT = urinary, psychosocial, organ specific, infection, neurologic or systemic, and tenderness; U-UDI = Urge-Urinary Distress Inventory; VAS = Visual Analogue Scale.

Table 2. Structural validity of the FSFI-19

Reference	Methodology	Outcome	Rating	Quality
Anis et al[16]	PCA	6-component structure	Indeterminate	Adequate
Bartula et al[19]	CFA	6 factors with item 14 removed: desire, arousal, lubrication, orgasm, satisfaction, pain	Sufficient	Very good
Bartula et al[92]	PCA	5 components: desire/arousal, lubrication, orgasm, satisfaction, pain	Indeterminate	Adequate
Baser et al[20]	EFA	5factors: desire/arousal, lubrication, orgasm, satisfaction, pain	Indeterminate	Adequate
Burri et al[93]	PCA	FSFI-19: Unrotated PCA identified five components. Although the sixth component had a considerably low eigenvalue, subsequent equamax rotation yielded the most consistent pattern of factor loadings using a 6-component structure. FSFI-LL: Unrotated PCA identified 3 components. Although the fourth and fifth components had considerably low eigenvalues, subsequent equamax rotation yielded the most consistent pattern of factor loadings using a 5-component structure.	Indeterminate	Adequate
Burri et al[93]	CFA	FSFI-19: The 6-factor solution was acceptable after allowing correlations between subscales and between a number of items. FSFI-LL: The 5-factor solution was acceptable after allowing correlations between subscales, and between a number of items.	Sufficient	Adequate
Burri et al[24]	PCA	Unrotated PCA identified 5 components with eigenvalues higher than 1. Although the sixth component had an eigenvalue <1, subsequent varimax rotation yielded the most consistent pattern of factor loadings using a 6-component structure.	Indeterminate	Adequate
Carpenter et al[26]	IRT	After pruning based on violations of local independence, on discrimination, and difficulty parameters; 9 items remained of the 19 items of the FSFI-19.	Insufficient	Adequate
Chang et al[27]	PCA	3 components were extracted and identified with eigenvalues greater than 1.03. These 3 components accounted for a total of 87.10% of the variance. Component 1, with an initial eigenvalue of 13.74, accounted for 72.32% of the explained variance. The 3 components were interpreted as "coitus" (15 items), "satisfaction" (2 items), and "desire" (2 items).	Indeterminate	Adequate
Fakhri et al[33]	PCA	The PCA yielded a best fitting, 5-component solution. All 5 components had eigenvalues of >1 and accounted for 70% of the total variance.	Indeterminate	Adequate
Fakhri et al[33]	CFA	6-factor structure showed inadequate fit ($\chi^2 = 826.60$; df = 136; GFI = 0.72; CFI = 0.81; NNFI = 0.63; SRMR, 0.18; PNFI = 0.63; RMSEA = 0.15). 5-factor structure showed acceptable fit ($\chi^2 = 304.07$; df = 142; GFI = 0.89; CFI = 0.95; NNFI = 0.94; SRMR, 0.08; PNFI = 0.71; RMSEA = 0.07).	Insufficient	Adequate
Forbes et al[38]	PCA	5 components with eigenvalues >1. The 5 components were clearly defined as desire and subjective arousal, lubrication, orgasm, satisfaction, and pain.	Indeterminate	Adequate
Ghassamia et al[39]	PCA	5 components were extracted with eigenvalues >1. The examination of the scree plot suggested that 4 or 5 dimensions underlie the FSFI. The components were interpreted as "Sexual Response" (11 items), "Sexual-related Pain" (3 items), "Sexual Desire" (2 items), "Sexual Satisfaction" (3 items).	Indeterminate	Adequate
Heng et al[41]	PCA	3 components were extracted with eigenvalues >1. The first component comprised sexual arousal, lubrication, and pain. The second component comprised orgasm and sexual satisfaction. Sexual desire alone made the third component.	Indeterminate	Adequate
Hevesi et al[44]	CFA	6-factor model had an acceptable fit (Satorra-Bentler $\chi^2 = 490.924$, df = 137, $P < .001$; χ^2/df = 3.583; CFI = 0.960; TLI = 0.950; RMSEA = 0.071; range = 0.065-0.078). However, most intercorrelations among the factors were very high. A bi-factor model where each item was associated with a general factor and with its domain-specific factor showed an improvement from the original model (Satorra-Bentler $\chi^2 = 272.630$, df = 123,	Sufficient	Very good

(continued)

Table 2. Continued

Reference	Methodology	Outcome	Rating	Quality
		$P < .001$; $\chi^2/df = 2.217$; CFI = 0.983; TLI = 0.976; RMSEA = 0.049; range = 0.041–0.057). It was found that in the total sample most observed variance was attributable to the general sexual function factor; whereas in the sexually active subsample most observed variance was attributable to the specific factors.		
Ismail et al[45]	PCA	Among the women without type 2 diabetes, 3 components were extracted with eigenvalues >1: Sexual desire/arousal, satisfaction, and pain. With the items in lubrication and orgasm domains loading on both satisfaction and pain. Among the women with type 2 diabetes, 3 components were extracted with eigenvalues >1: A component comprising of lubrication, orgasm, and pain; satisfaction, and desire/arousal.	Indeterminate	Adequate
Kalmbach et al[47]	CFA	Bad model fit of six-factor model (χ^2 [137] = 683.28, $P < .001$, CFI = .91, TLI = .88, RMSEA = .07). Adding latent variables describing whether an item was positively or negatively worded increased the fit (χ^2 [118] = 303.01, $P < .001$, CFI = .97, TLI=.95, RMSEA=.04).	Insufficient	Very good
Liu et al[50]	PCA	5 components were extracted with an eigenvalue >1, accounting for 77.57% of the total variance. The first component consisted of a mixture of desire/arousal, and the rest were lubrication, orgasm, satisfaction, and pain.	Indeterminate	Adequate
Liu et al[50]	CFA	A 6-factor model showed a bad fit (CMIN/DF = 3.12, GFI = .83, CFI = .91, RMSEA = .100). A 5-factor model showed an acceptable fit (CMIN/DF = 3.08, CFI = .91, GFI = .83, RMSEA = .099). The 5 factors included desire/arousal, lubrication, orgasm, satisfaction, and pain.	Insufficient	Adequate
Nowosielski et al[58]	PCA	5 components were extracted, accounting for 83.62% of the total variance. The components reflected desire/arousal, lubrication, orgasm, satisfaction, and pain.	Indeterminate	Adequate
Opperman et al[59]	CFA	A first-order, 1 -factor model showed a bad fit: χ^2 (152, N = 85) = 664.45, $P < .001$ ($\chi^2/df = 4.4$; GFI = .55, TLI = .34, CFI = .41, and RMSEA = .20). A first-order, 6-factor model with correlations among factors showed a good fit: χ^2 (137, N = 85) = 178.96, $P = .009$ ($\chi^2/df = 1.3$; GFI = .83, TLI = .94, CFI = .95, and RMSEA = .06). A second-order, 6-factor model with 1 second-order factor showed a decrement in fit compared to the first-order, 6-factor model: χ^2 (146, N = 85) = 199.72, $P = .002$ ($\chi^2/df = 1.4$; GFI = .80, TLI = .93, CFI = .94, and RMSEA = .066). A first-order, 5-factor model showed a slight decrement in fit compared with the first-order and second-order, 6-factor models: χ^2 (137, N = 85) = 215.89, $P < .001$ ($\chi^2/df = 1.5$; GFI = .79, TLI = .90, CFI = .92, and RMSEA = .079). Delta χ^2 tests of differences indicated a better fit (Delta χ^2 [9, n = 85] = 21.62, $P < .05$) of the first-order, 6-factor model (χ^2 [137, N = 85] = 178.96) vs the second-order, 6-factor model (χ^2 [146, N = 85] = 199.72), as well as a significantly better fit (Delta χ^2 [4, N = 85] = 37.79, $P < .01$) vs the first-order, 5-factor model (χ^2 [142, N = 85] = 215.89).	Sufficient	Inadequate
Rehman et al[63]	PCA	5-component structure: Desire, arousal, lubrication, orgasm, satisfaction and pain with eigenvalues 7.556, 3.457, 2.939, 2.926 and 0.633, respectively. These 5 components accounted for 92.164% of the explained variance.	Indeterminate	Adequate
Rillon-Tabil et al[65]	PCA	6-component structure	Indeterminate	Inadequate
Rosen et al[2]	CFA	5 factors: Desire/arousal, lubrication, orgasm, satisfaction, and pain; but desire/arousal was split into 2 factors due to theoretical considerations	Indeterminate	Very good
Rosen et al[2]	PCA	5 components: Desire/arousal, lubrication, orgasm, satisfaction, and pain	Indeterminate	Inadequate
Sun et al[73]	PCA		Indeterminate	Adequate

(continued)

Table 2. Continued

Reference	Methodology	Outcome	Rating	Quality
		4 components were extracted with eigenvalue >1. The first component was a mixture of arousal/orgasm/satisfaction and the remaining 3 components were lubrication, pain, and desire. These 4 components accounted for a total of 75.01% of the explained variance		
Takahashi et al[74]	EFA	5-factor structure found: desire/arousal, lubrication, orgasm, satisfaction, and pain	Insufficient	Adequate
Ter Kuile et al[75]	SCA	6-component structure explained 88.6% variance.	Indeterminate	Adequate
Ter Kuile et al[75]	PCA	6-component structure explained 81.6% variance.	Indeterminate	Adequate
Vallejo-Medina et al[78]	EFA	5-factor structure found with a clear Arousal—Desire fusion.	Insufficient	Adequate
Vallejo-Medina et al[78]	CFA	A 6-factor uncorrelated factor model showed a bad fit (S-B χ^2 [df = 146] = 550.02, $P < .001$, CFI = .976, RMSEA =.076, AIC = 258.02). A 6-factor correlated factor model showed a good fit (S-B χ^2 [df = 131] = 209.31, $P <$. 001, CFI = .995, RMSEA = .036, AIC = −52.68). A second-order, 6-factor model showed a good fit (S-B χ^2 [df = 145] = 353.60, $P < .001$, CFI = .988, RMSEA = .055, AIC = 63.60). A 5-factor correlated model showed a good fit (S-B χ^2 [df = 137] = 338.86, $P < .001$, CFI = .988, RMSEA =.056, AIC = 64.86).	Sufficient	Adequate
Wiegel et al[82]	PCA	A PCA on a sample of sexually functional and dysfunctional women (n = 272) showed a 5-component structure: desire/arousal, lubrication, orgasm, pain, and satisfaction. A PCA in women with sexual dysfunction resulted in 4 components with eigenvalues >1 and 1 component with eigenvalue of 0.98. The 4 components were interpreted as pain, orgasm, lubrication, desire/arousal/satisfaction. When taking the fifth component into account, they were interpreted as desire/arousal, lubrication, orgasm, pain, and satisfaction. A PCA in women without sexual dysfunction resulted in 5 components with eigenvalues >1, which were interpreted as desire/arousal, orgasm/arousal, lubrication, satisfaction, and pain. A PCA of the combined group (dysfunctional and non-dysfunctional; n = 527) resulted in 5 components with eigenvalues >1, which were interpreted as: desire/arousal, orgasm/arousal, lubrication, satisfaction, and pain.	Indeterminate	Adequate
Witting et al[83]	EFA	4 factors had an eigenvalue > 1. The fifth factor had an eigenvalue of 0.84 for Twin 1 group and 0.85 for Twin 2 group. The sixth factor had an eigenvalue of 0.57 and 0.62, respectively. After exploring 4-, 5-, and 6-factor solutions, it was decided to use the 6-factor solution due to interpretability. The 6-factor solution explained 76.6% of the variance for Twin 1, whereas a general factor only explained 48.6%. The corresponding figures for Twin 2 were 75.3% and 47.0%, respectively. This suggested that a 1-factor model was not adequate.	Indeterminate	Adequate
Witting et al[83]	CFA	A six-factor model in the Twin 1 group showed a bad fit (χ^2 [df = 137] = 789.03, GFI = .924; NFI = .956, RMSEA = .067, AIC 895.08). The results for Twin 2 were similar.	Insufficient	Adequate
Wolpe et al[84]	PCA	2-component structure was found with the first component explaining 76.66% of variance, and the second component explaining 6.16% of variance.	Indeterminate	Adequate
Wylomanski et al[85]	EFA	6-factor structure was found, explaining 71.4% of variance.	Indeterminate	Adequate
Wylomanski et al[85]	CFA	6-factor model did not fit the data. The model was adjusted based on modification indices, adding covariance between error terms of 4-item pairs: 7—10, 15—16, 3—4 and 8—9. Each pair of items included a similar content. This adjusted model showed a good fit (Q = 2.8, CFI = 0.98, RMSEA = 0.06 and SRMR = 0.03).	Insufficient	Adequate

AIC = akaike information criterion; CFA = confirmatory factor analysis; CFI = comparative fit index; EFA = exploratory factor analysis; FSFI = Female Sexual Function Index; GFI = goodness of fit index; NFI = normed fit index; NNFI = non-normed fit index; PCA = principal component analysis; PNFI = parsimony normed fit index; RMSEA = root mean square error of approximation; S-B = Satorra-Bentler; SRMR = standardized root mean square residual; TLI = Tucker Lewis index.

Table 3. Test-retest reliability of the FSFI

Reference	Coefficient	Total score	DE	AR	LU	OR	SA	PA	Rating	Quality
FSFI-19										
Anis et al[16]	Correlation	.98	.92	.98	.97	.98	.96	.97	Sufficient	Doubtful
Bartula et al[19]	Correlation		.86	.82	.78	.80	.76	.75	Sufficient	Doubtful
Borello-France et al[22]	ICC	.91	.84	.86	.82	.90	.79	.88	Sufficient	Adequate
Chang et al[27]	Correlation	.69							Insufficient	Doubtful
Fakhri et al. (2012) [33]	ICC	.77	.84	.78	.86	.82	.79	.73	Sufficient	Adequate
Filocamo et al[36]	Correlation	.95	.93	.93	.95	.92	.92	.93	Sufficient	Doubtful
Ghassamia et al[39]	Correlation	.82	.66				.72	.78	Sufficient	Doubtful
Liu et al[50]	Correlation	.84		.68	.83				Sufficient	Doubtful
Nowosielski et al[58]	ICC day 7	.83	.83	.89	.85	.88	.87	.80	Sufficient	Adequate
	ICC day 28	.75	.80	.86	.80	.81	.78	.73		
Rehman et al[63]	ICC	.99	1.00	1.00	.99	.98	.99	1.00	Sufficient	Doubtful
Rillon-Tabil et al[65]	Correlation	.99							Sufficient	Doubtful
Rosen et al[2]	Correlation	.88	.83	.85	.86	.80	.83	.79	Sufficient	Doubtful
Ryding et al[68]	Correlation	.77 - .95	.67 -.89	.62 - .90	.35 - .85	.65 - .86	.65 - .86	.10 - .90	Insufficient	Doubtful
Sidi et al[70]	Correlation		.87	.77	.95	.97	.95	.86	Sufficient	Doubtful
Sun et al[73]	Correlation	.80 - .86	.72 - .85	.78 - .95	.74 - .93	.85 - .89	.80 - .86	.69 - .90	Sufficient	Doubtful
Takahashi et al[74]	ICC		.73 — 1.00	.73 — 1.00	.73 — 1.00	.73 — 1.00	.73 — 1.00	.73 — 1.00	Sufficient	Inadequate
Ter Kuile et al[75]	Correlation	.93	.72	.85	.82	.71	.90	.97	Sufficient	Inadequate
Verit et al[80]	Correlation	.90—.92	.79 - .81	.85 - .87	.85 - .88	.83 - .87	.83 - .85	.89	Sufficient	Doubtful
Wolpe et al[84]	ICC	.94							Sufficient	Adequate
Wylomanski et al[85]	ICC	.99	.97	.99	.97	.96	.89	.99	Sufficient	Adequate
Zachariou et al[86]	Correlation	.91							Sufficient	Doubtful
FSFI-6										
Isidori et al[4]	Correlation	.95							Sufficient	Doubtful
Lee et al[89]	ICC	.61							Insufficient	Adequate
FSFI-BC (34 items)										
Bartula et al[92]	Correlation		.72 - .88		.71 - .72	.63 - .85	.86	.77 - .80	Sufficient	Doubtful

AR = arousal; DE = desire; FSFI = Female Sexual Function Index; ICC = intraclass correlation; LU = lubrication; OR = orgasm; PA = pain; SA = satisfaction.

Table 4. Criterion validity

Reference	Instrument	AUC	Cutoff	Sensitivity	Specificity	PPV	PNV	Rating	Quality
FSFI-19									
Anis et al[16]	FSD diagnosis	.99		.97	.93			Sufficient	Very good
Fakhri et al[33]	FSD diagnosis	.91		.82	.86			Sufficient	Very good
Gerstenberger et al[40]	HSDD diagnosis			.70 - .97	.84 - .97			Sufficient	Very good
Ma et al[51]	FSD diagnosis Total FSFI		23.45	.67	.73			Sufficient	Very good
	Low desire	.73	<2.8	.55	.78				
	Arousal disorder	.74	<3.16	.62	.77				
	Lubrication disorder	.85	<4.06	.86	.70				
	Orgasm disorder	.85	<3.9	.83	.74				
	Sexual pain	.79	<3.9	.65	.81				
Nowosielski et al[58]	FSD diagnosis	.93		.87	.83	.86		Sufficient	Very good
Ryding et al[68]	HSDD diagnosis			.96	.97			Indeterminate	Very good
Sidi et al[70]	FSD diagnosis	.99		.99	.97			Sufficient	Very good
Ter Kuile et al[75]	FSD diagnosis	.98		.95	.92	.96	.89	Sufficient	Adequate
Wiegel et al[82]	FSD diagnosis	.90	26.55	.88 - .89	.71 - .73			Sufficient	Very good
Zachariou et al[86]	FSD diagnosis	.86		.72	.93			Sufficient	Doubtful
FSFI-6									
Isidori et al[4]	FSD diagnosis	.98	19	.96	.91	.95	.93	Sufficient	Very good
Lee et al[89]	FSD diagnosis	.95		.90	.86			Sufficient	Very good

AUC = area under the curve; FSD = female sexual disorder; FSFI = Female Sexual Function Index; HSDD = hypoactive sexual desire disorder.

Methodologic quality was rated as "adequate"[14,19,48,92] or "doubtful".[2,58,68] The "doubtful" ratings were due to lack of information on the measurement properties of the comparator instrument.

2 studies of "adequate" quality[14,19] and 3 studies of "doubtful" quality[2,68,76] reported low correlation coefficients that were rated as sufficient. 2 studies of "adequate" quality,[48,92] and 1 study of "doubtful" quality [58] reported multiple correlation coefficients >0.30 and were rated as insufficient. None of the studies that investigated the FSFI-6[4,88–91] reported on divergent validity.

Criterion Validity

10 studies reported on criterion validity of the FSFI-19 using the gold standard of FSD or HSDD diagnosis[16,33,40,51,58,68,70,75,82,86] (Table 4). Methodologic quality was rated as "very good",[16,33,40,51,58,68,70,82] as "adequate",[75] or as "doubtful".[86] The "doubtful" rating was due to a small sample size ("other flaws" in COSMIN methodologic quality).[86] 1 study did not report an area under the curve (AUC) and was thus rated indeterminate.[68] All remaining studies reported AUC values that were rated as sufficient.

2 studies reported on criterion validity of the FSFI-6 using FSD diagnosis as the gold standard[4,89] (Table 4). Methodologic quality was rated as "very good" for both studies. Both studies reported AUC values that were rated as sufficient.

Data Synthesis

The synthesized ratings of the measurement properties across all studies can be found in Table 5. The evidence of structural validity of the FSFI-19 was rated as inconsistent, because 6-factor, 5-factor, and other factor structures were reported. The evidence was evaluated as low quality because of this inconsistency, as well as a risk of bias because many studies reported a principle component analysis (PCA) instead of an exploratory or confirmatory factor analysis. The evidence of internal consistency of the FSFI-19 was rated as sufficient but of moderate quality, due to 15.8% (n = 6) of studies reporting insufficient internal consistency.

The evidence of test-retest reliability of the FSFI-19 was rated as sufficient but of low quality, due to 13.0% (n = 3) of studies reporting insufficient test-retest reliability, as well as risk of bias as many studies reported Pearson correlation instead of the intraclass correlation coefficient. The use of Pearson correlations are problematic, because they do not control for systematic error variance, which is a product of measuring the same individual twice. The intraclass correlation coefficient controls for this systematic error variance, and without this control, test-retest reliability may be overestimated.[95,96] The evidence of construct validity (hypothesis testing) of the FSFI-19 was rated as inconsistent with moderate quality, because 28.6% (n = 18) of studies reported insufficient values. The evidence of criterion validity of the FSFI-19 was rated as sufficient with high quality. The evidence of measurement error and responsiveness of the FSFI-19 were rated as indeterminate, because no data were reported on these measurement properties.

Evidence of structural validity of the FSFI-6 was rated as indeterminate because no data were reported on this measurement property. The evidence of internal consistency of the FSFI-6 was rated as indeterminate, as evidence for unidimensionality was

Table 5. Ratings of measurement properties

Measurement property	Rating of measurement property	Quality of evidence
FSFI-19		
Structural validity	Inconsistent	Low
Internal consistency	Sufficient	Moderate
Reliability	Sufficient	Low
Measurement error	Indeterminate	
Construct validity	Inconsistent	Moderate
Criterion validity	Sufficient	High
Responsiveness	Indeterminate	
FSFI-6		
Structural validity	Indeterminate	
Internal consistency	Indeterminate	
Reliability	Inconsistent	Low
Measurement error	Indeterminate	
Construct validity	Inconsistent	Very low
Criterion validity	Sufficient	Moderate
Responsiveness	Indeterminate	

FSFI = Female Sexual Function Index.

missing. The evidence of line number test-retest reliability of the FSFI-6 was rated as inconsistent and of low quality, due to risk of bias because only 2 studies reported on test-retest reliability, of which 1 study was of doubtful methodologic quality. The evidence of construct validity (hypothesis testing) of the FSFI-6 was rated as inconsistent with very low quality, because there were as many studies reporting sufficient (50%) as insufficient values (50%), as well as a risk of bias due to methodologic quality of the studies. The evidence of criterion validity of the FSFI-6 was rated as sufficient and of moderate quality, as the evidence was based on only 2 studies. The evidence of measurement error and responsiveness of the FSFI-6 were rated as indeterminate, because no data were reported on these measurement properties.

DISCUSSION

This systematic review investigated the evidence of the measurement properties of the FSFI-19[2] and FSFI-6.[4] Concerning the FSFI-19, the evidence on internal consistency was sufficient and of moderate quality. The evidence on test-retest reliability was also sufficient but of low quality due to some inconsistencies and many studies not using the intraclass correlation. The evidence on criterion validity was also sufficient and of high quality. The evidence on structural validity was inconsistent and of low quality. Studies found either evidence for the theorized 6-factor structure, a 5-factor structure (with a merging of desire and arousal) or structures with <5 factors. Evidence on construct validity was inconsistent and of moderate quality. No data was found on measurement error and responsiveness.

Concerning the FSFI-6, the evidence on criterion validity was sufficient and of moderate quality. Evidence on reliability was inconsistent with low quality of evidence, due to a high risk of bias. Evidence on construct validity was inconsistent with very low quality of evidence, due to as many studies reporting sufficient as insufficient values as well as a risk of bias due to methodologic quality. The evidence of structural validity, internal consistency, measurement error, and responsiveness were rated as indeterminate.

Regarding the structural validity of the FSFI-19, there was more evidence against than in favor of the hypothesized 6-factor structure. This is in line with the revisions made in the DSM-5 to the model of FSD.[5] Whereas only 2 articles showed direct support for a 5-factor structure,[33,50] 12 more studies showed indirect support for a 5-factor solution through use of PCAs.[2,20,24,33,38,39,50,58,63,82,92,93] However, other PCAs resulted in other factor structures: 4 factors,[73,82,83] 3 factors,[27,41,45] or 2 factors.[84] Based on the wide range of reported factor structures, it may be that the factor structure of the FSFI-19 is different for different subgroups or nationalities. In fact, some studies investigated factor structures in subgroups, and results suggest that there are different factor structures of the FSFI-19 among women with FSD and women without FSD.[44,82] Unfortunately, neither study performed a test of measurement invariance, and, as such, there is no direct evidence for this hypothesis. Such differences in factor structure in different subgroups may be related to a number of theoretical positions. For one, differing motivations for sex for women with arousal disorder vs without arousal disorder [97] may suggest that arousal and desire may be a singular motivation (ie, a singular construct) for women without arousal disorder, but not for women with an arousal disorder. Furthermore, the position that FSD represents a spectrum of disorders with extensive overlap[98–100] implies that the constructs measured by the FSFI-19 may be different for women suffering from differing (yet overlapping) sexual disorders.

Nevertheless, based on this systematic review, we conclude that the use of the 6 subscales may not be valid in all patient groups. Instead, we see compelling evidence to merge the subscales of arousal and desire. For confirming whether these subscales should be merged, and whether the constructs measured by the FSFI-19 are different for subpopulations of women, a large-scale validation study focusing on testing measurement invariance across patient subgroups and nationalities, as well as multiple factor structures through use of confirmatory factor analyses is needed. In the meantime we recommend that, to ensure valid interpretation of their results, researchers perform and report on structural validity of the FSFI-19 when presenting the results of their studies.

Although we rated the evidence on internal consistency as sufficient for the FSFI-19, it needs to be noted that unidimensionality of the subscales is a prerequisite or interpreting their internal consistency. As the structural validity of the FSFI-19 is shaky at best, our rating of the evidence on internal consistency is mostly intended as an interpretation for the subscales that are

found as being unidimensional in most analyses: lubrication, satisfaction, and pain.

With respect to the structural validity of the FSFI-6, validation studies investigating the unidimensional nature of the instrument are of importance, because no studies investigated this measurement property. The FSFI-6 results in 1 score representing FSD in general, and it is crucial to determine whether all items represent the construct of FSD in a unidimensional manner. However, structural validity of the FSFI-6 is likely not so straightforward: because the 6 best-performing items of the FSFI-19 were selected from each domain, any and all overlap of constructs of the subscales of the FSFI-19 will also be represented in the FSFI-6. Furthermore, as the FSFI-6 is a composite of multiple constructs, it is unlikely to be unidimensional because it is based on a formative model instead of a reflective model. This raises issues with the interpretation of the FSFI-6 total score, because it may not reflect 1 general construct of FSD. For a total score of a multidimensional instrument, it is unclear what exact construct is represented by the total score.

The evidence on internal consistency of the FSFI-6 cannot yet be determined, because the unidimensionality (a prerequisite for internal consistency) has not yet been determined. However, if the FSFI-6 is found to be unidimensional, internal consistency is likely to be rated as sufficient, because 3 studies of very good quality found values of Cronbach's α that would be rated as sufficient.

Research into measurement error and responsiveness is necessary as well. With the high use of the FSFI-19 and FSFI-6 in clinical practice and clinical research, it is of importance to know which change can be distinguished from measurement error. To further the knowledge, we recommend that researchers perform test-retest reliability studies to calculate the limits of agreement or smallest detectable change.[95] Furthermore, an anchor-based study is recommended to determine the minimal important change to be able to interpret the limits of agreement or smallest detectable change.[101]

Combining the concerns surrounding structural validity, inconsistent findings on multiple measurement properties, the low quality of many of the included studies, and the missing information on multiple measurement properties; questions are raised on the validity and reliability of the FSFI-19 and FSFI-6 as measures of FSD. The content validity of the FSFI-19 has been challenged previously,[102] where the FSFI-19 was described as a measure of vaginal intercourse, and not FSD. Combining concerns regarding content validity, as well as our concerns regarding structural validity, it is unclear whether the FSFI measures FSD, or a selection of symptoms related to FSD. Regardless of these concerns, evidence for criterion validity is strong. Pragmatically, the FSFI is a good screening tool for the current definition of FSD. However, from a psychometric point of view, the above concerns are serious. Given the high frequency of use of both the FSFI-19 and the FSFI-6 in clinical screening for FSD, as well as an outcome measure for clinical trials, it is of importance that more research is performed into the measurement properties and content validity.

A limitation of this review is that we did not investigate content validity. Content validity needs to be established before other measurement properties can be evaluated.[6] A future investigation of content validity is warranted. Another limitation of this review is the use of a precise rather than a sensitive search filter regarding measurement properties. The sensitive filter was developed to capture every relevant hit, at the expense of capturing more false-positive search hits. Meanwhile the specific filter was developed to capture as many relevant hits, while decreasing the number of false-positive search hits. The sensitivity of the precise filter was 93% in a random set of PubMed records, whereas the sensitivity of the sensitive search filter was 97%.[12] The use of the precise filter was a pragmatic choice over the available sensitive filter because the initial search encompassed 39 PROMs (including the FSFI-19 and FSFI-6), and the sensitive filter would provide too many hits for feasible screening. The possibility remains that the precise filter missed some validation studies of the FSFI-19 and FSFI-6.

CONCLUSIONS

Based on this systematic review, we conclude that with respect to internal consistency, reliability, and criterion validity, the FSFI-19 meets psychometric criteria, but has not been shown to meet psychometric criteria for structural validity, measurement error, construct validity, and responsiveness. Evidence on structural validity suggests a merging of the subscales arousal and desire. Such a merging of subscales has consequences for the interpretation of the FSFI-19 in both clinical practice and research. To investigate this possible adjustment to the FSFI-19, as well as the suggestion that factor structures may be population-dependent; a large-scale cross-cultural study design or an individual patient data meta-analysis, applying Confirmatory Factor Analysis, measurement invariance tests, and calculating the limits of agreement and smallest detectable change, is recommended.

The FSFI-6 meets psychometric criteria with respect to criterion validity. Structural validity, internal consistency, reliability, measurement error, construct validity, and responsiveness require further research. Most importantly for future research is determining the unidimensionality of the FSFI-6. Regardless of these concerns, evidence for criterion validity is strong for both the FSFI-19 and FSFI-6, and pragmatically, they are good screening tools for the current definition of FSD.

ACKNOWLEDGMENTS

We thank Margot Veeger for help with the screening procedures. We thank Anja van der Hout, Heleen Melissant, and Evalien Veldhuijzen for help with the data extraction procedures.

Corresponding Author: Irma M. Verdonck-de Leeuw, PhD, Vrije Universiteit Amsterdam, Department of Clinical, Neuro- and

Developmental Psychology, The Netherlands. Tel: +31 20 444 0931; Fax: +31 20 444 3688; E-mail: IM.Verdonck@vumc.nl

Conflicts of Interest: The authors report no conflicts of interest.

Funding: This work was supported by the Dutch Cancer Society [grant number VUP 2014-7202].

STATEMENT OF AUTHORSHIP

Category 1

(a) Conception and Design
Koen I. Neijenhuijs; Neil K. Aaronson; Mogens Groenvold; Bernhard Holzner; Caroline B. Terwee; Pim Cuijpers; Irma M. Verdonck-de Leeuw

(b) Acquisition of Data
Koen I. Neijenhuijs; Nienke Hooghiemstra; Karen Holtmaat

(c) Analysis and Interpretation of Data
Koen I. Neijenhuijs; Nienke Hooghiemstra; Karen Holtmaat

Category 2

(a) Drafting the Article
Koen I. Neijenhuijs

(b) Revising It for Intellectual Content
Nienke Hooghiemstra; Karen Holtmaat; Neil K. Aaronson; Mogens Groenvold; Bernhard Holzner; Caroline B. Terwee; Pim Cuijpers; Irma M. Verdonck-de Leeuw

Category 3

(a) Final Approval of the Completed Article
Koen I. Neijenhuijs; Nienke Hooghiemstra; Karen Holtmaat; Neil K. Aaronson; Mogens Groenvold; Bernhard Holzner; Caroline B. Terwee; Pim Cuijpers; Irma M. Verdonck-de Leeuw

REFERENCES

1. Basson R, Berman J, Burnett A, et al. Report of the international consensus development conference on female sexual dysfunction: Definitions and classifications. **J Urol 2000;163:888-893.**
2. Rosen R, Brown C, Heiman J, et al. The Female Sexual Function Index (FSFI): A multidimensional self-report instrument for the assessment of female sexual function. **J Sex Marital Ther 2000;26:191-208.**
3. Locke HJ, Wallace KM. Short marital-adjustment and prediction tests: Their reliability and validity. **Marriage and Family Living 1959;21:251-255.**
4. Isidori AM, Pozza C, Esposito K, et al. Development and validation of a 6-item version of the female sexual function index (FSFI) as a diagnostic tool for female sexual dysfunction. **J Sex Med 2010;7:1139-1146.**
5. American Psychiatric Association. DSM-5 Task Force. Diagnostic and Statistical Manual of Mental Disorders: DSM-5. Philadelphia: American Psychiatric Association; 2013.
6. Prinsen CAC, Mokkink LB, Bouter LM, et al. COSMIN guideline for systematic reviews of patient-reported outcome measures. **Qual Life Res 2018;27:1147-1157.**
7. Mokkink LB, Terwee CB, Patrick DL, et al. The COSMIN study reached international consensus on taxonomy, terminology, and definitions of measurement properties for health-related patient-reported outcomes. **J Clin Epidemiol 2010;63:737-745.**
8. van der Hout A, van Uden-Kraan CF, Witte BI, et al. Efficacy, cost-utility, and reach of an eHealth self-management application "Oncokompas" that facilitates cancer survivors to obtain optimal supportive care: Study protocol for a randomized controlled trial. **Trials 2017;18:228.**
9. Lubberding S, van Uden-Kraan CF, Te Velde EA, et al. Improving access to supportive cancer care through an eHealth application: A qualitative needs assessment among cancer survivors. **J Clin Nurs 2015;24:1367-1379.**
10. Jansen F, van Uden-Kraan CF, Van Zwieten V, et al. Cancer survivors' perceived need for supportive care and their attitude towards self-management and eHealth. **Support Care Cancer 2015;23:1679-1688.**
11. Duman-Lubberding S, van Uden-Kraan CF, Jansen F, et al. Feasibility of an eHealth application "OncoKompas" to improve personalized survivorship cancer care. **Supportive Care Cancer 2016;24:2163-2171.**
12. Terwee CB, Jansma EP, Riphagen II, et al. Development of a methodological PubMed search filter for finding studies on measurement properties of measurement instruments. **Qual Life Res 2009;18:1115-1123.**
13. Mokkink LB, de Vet HCW, Prinsen CAC, et al. COSMIN Risk of Bias checklist for systematic reviews of Patient-Reported Outcome Measures. **Qual Life Res 2018;27:1171-1179.**
14. Achimas-Cadariu P, Irimie A, Iancu M, et al. Identification and validation of quality of life measures in a population of women with premalignant and malignant pathology at childbearing age. **J Cogn Behav Psychother 2013;13:409-420.**
15. Ahmed MR, Shaaban MM, Meky HK. Assessment of sexually related personal distress accompanying premenopausal sexual dysfunction with an Arabic version of the Female Sexual Distress Scale. **Int J Gynecol Obstet 2017;139:65-70.**
16. Anis T, Gheit AS. Arabic translation of Female Sexual Function Index and validation in an Egyptian population. **J Sex Med 2011;8:3370-3378.**
17. Aydin S, Onaran OI, Topalan K, et al. Development and validation of Turkish version of The Female Sexual Distress Scale—Revised. **Sex Med 2016;4:E43-E50.**
18. Azimi Nekoo E, Burri A, Ashrafti F, et al. Psychometric properties of the Iranian version of the Female Sexual Distress Scale-Revised in women. **J Sex Med 2014;11:995-1004.**
19. Bartula I, Sherman KA. The Female Sexual Functioning Index (FSFI): Evaluation of acceptability, reliability, and validity in women with breast cancer. **Support Care Cancer 2015;23:2633-2641.**
20. Baser RE, Li YL, Carter J. Psychometric validation of the female sexual function index (FSFI) in cancer survivors. **Cancer 2012;118:4606-4618.**
21. Bloemendaal LBA, Laan ETM. The psychometric properties of the Sexual Excitation/Sexual Inhibition Inventory for Women (SESII-W) within a Dutch population. **J Sex Res 2015;52:69-82.**

22. Borello-France D, Dusi J, O'Leary M, et al. Test-retest reliability of the Urge-Urinary Distress Inventory and Female Sexual Function Index in women with multiple sclerosis. **Urol Nurs 2008;28:30-35.**
23. Bornefeld-Ettmann P, Steil R, Hoefling V, et al. Validation of the German version of the Sexual Self-Esteem Inventory for Women and its application in a sample of sexually and physically abused women. **Sex Roles 2018;79:109-122.**
24. Burri A, Porst H. Preliminary validation of a German version of the Sexual Complaints Screener for Women in a female population sample. **Sex Med 2018;6:123-130.**
25. Carpenter JS, Reed SD, Guthrie KA, et al. Using an FSDS-R item to screen for sexually related distress: A MsFLASH Analysis. **Sex Med 2015;3:7-13.**
26. Carpenter JS, Jones SMW, Studts CR, et al. Female Sexual Function Index Short Version: A MsFLASH Item Response Analysis. **Arch Sex Behav 2016;45:1897-1905.**
27. Chang S-R, Chang T-C, Chen K-H, et al. Developing and validating a Taiwan version of the Female Sexual Function Index for pregnant women. **J Sex Med 2009;6:1609-1616.**
28. Clayton A, Segraves R, Leiblum S, et al. Reliability and validity of the Sexual Interest and Desire Inventory-Female (SIDI-F), a scale designed to measure severity of female hypoactive sexual desire disorder. **J Sex Marital Ther 2006;32:115-135.**
29. Clayton A, Goldmeier D, Nappi R, et al. Validation of the Sexual Interest And Desire Inventory-Female in hypoactive sexual desire disorder. **J Sex Med 2010;7:3918-3928.**
30. Constantine ML, Pauls RN, Rogers RR, et al. Validation of a single summary score for the Prolapse/Incontinence Sexual Questionnaire-IUGA revised (PISQ-IR). **Int Urogynecol J 2017;28:1901-1907.**
31. DeRogatis LR, Allgood A, Auerbach P, et al. Validation of a Women's Sexual Interest Diagnostic Interview—Short Form (WSID-SF) and a Daily Log of Sexual Activities (DLSA) in postmenopausal women with hypoactive sexual desire disorder. **J Sex Med 2010;7:917-927.**
32. Eaton A, Baser R, Seidel B, et al. Validation of clinical tools for vaginal and vulvar symptom assessment in cancer patients and survivors. **J Sex Med 2017;14:144-151.**
33. Fakhri A, Pakpour A, Burri A, et al. The Female Sexual Function Index: Translation and validation of an Iranian version. **J Sex Med 2012;9:514-523.**
34. Farkas B, Tiringer I, Farkas N, et al. Hungarian language validation of the Pelvic Organ Prolapse/Incontinence Sexual Questionnaire, IUGA-Revised (PISQ-IR). **Int Urogynecol J 2016;27:1831-1836.**
35. Ferguson SE, Urowitz S, Massey C, et al. Confirmatory factor analysis of the sexual adjustment and body image scale in women with gynecologic cancer. **Cancer 2012;118:3095-3104.**
36. Filocamo MT, Serati M, Marzi VL, et al. The Female Sexual Function Index (FSFI): Linguistic validation of the Italian version. **J Sex Med 2014;11:447-453.**
37. Flynn KE, Reeve BB, Lin L, et al. Construct validity of the PROMIS sexual function and satisfaction measures in patients with cancer. **Health Qual Life Outcomes 2013;11:1.**
38. Forbes MK, Baillie AJ, Schniering CA. Critical flaws in the Female Sexual Function Index and the International Index of Erectile Function. **J Sex Res 2014;51:485-491.**
39. Ghassamia M, Asghari A, Shaeiri MR, et al. Validation of psychometric properties of the Persian version of the Female Sexual Function Index. **Urol J 2013;10:878-885.**
40. Gerstenberger E, Rosen R, Brewer J, et al. Sexual Desire and the Female Sexual Function Index (FSFI): A sexual desire cutpoint for clinical interpretation of the FSFI in women with and without hypoactive sexual desire disorder. **J Sex Med 2010;7:3096-3103.**
41. Heng YS, Sidi H, Jaafar NRN, et al. Phases of female sexual response cycle among Malaysian women with Infertility: A factor analysis study. **Asia-Pacific Psychiatry 2013;5:50-54.**
42. Herbenick D, Reece M. Development and validation of the female genital self-image scale. **J Sex Med 2010;7:1822-1830.**
43. Herbenick D, Schick V, Reece M, et al. The Female Genital Self-Image Scale (FGSIS): Results from a Nationally Representative Probability Sample of Women in the United States. **J Sex Med 2011;8:158-166.**
44. Hevesi K, Meszaros V, Kovi Z, et al. Different characteristics of the Female Sexual Function Index in a sample of sexually active and inactive women. **J Sex Med 2017;14:1133-1141.**
45. Ismail A, Bau R, Sidi H, et al. Factor analysis study on sexual responses in women with Type 2 diabetes mellitus. **Comprehensive Psychiatr 2014;55:S34-S37.**
46. Jing L-W, Zhang C, Jin F, et al. Development of a quality of sexual life questionnaire for breast cancer survivors in Mainland China. **Med Sci Monitor 2018;24:4101-4112.**
47. Kalmbach D, Ciesla J, Janata J, et al. The validation of the Female Sexual Function Index, Male Sexual Function Index, and Profile of Female Sexual Function for use in healthy young adults. **Arch Sex Behav 2015;44:1651-1662.**
48. Likes WM, Stegbauer C, Hathaway D, et al. Use of the female sexual function index in women with vulvar intraepithelial neoplasia. **J Sex Marital Ther 2006;32:255-266.**
49. Liu B, Su M, Zhan H, et al. Adding a sexual dysfunction domain to UPOINT system improves association with symptoms in women with interstitial cystitis and bladder pain syndrome. **Urology 2014;84:1308-1313.**
50. Liu H, Yu J, Chen Y, et al. Sexual function in cervical cancer patients: Psychometric properties and performance of a Chinese version of the Female Sexual Function Index. **Eur J Oncol Nurs 2016;20:24-30.**
51. Ma J, Pan L, Lei Y, et al. Prevalence of female sexual dysfunction in urban Chinese women based on cutoff scores of the Chinese version of the Female Sexual Function Index: A preliminary study. **J Sex Med 2014;11:909-919.**
52. Meston C. Validation of the Female Sexual Function Index (FSFI) in women with female orgasmic disorder and in women with hypoactive sexual desire disorder. **J Sex Marital Ther 2003;29:39-46.**
53. Meston C, Trapnell P. Development and validation of a five-factor sexual satisfaction and distress scale for women: The

Sexual Satisfaction Scale for Women (SSS-W). **J Sex Med 2005;2:66-81.**

54. Mestre M, Lleberia J, Pubill J, et al. Spanish version of the Pelvic Organ Prolapse/Urinary Incontinence Sexual Questionnaire IUGA-Revised (PISQ-IR): Transcultural validation. **Int Urogynecol J 2017;28:1865-1873.**
55. Mohammadi K, Rahnama P, Montazeri A, et al. The Multiple Sclerosis Intimacy and Sexuality Questionnaire-19: Reliability, validity, and factor structure of the Persian version. **J Sex Med 2014;11:2225-2231.**
56. Mohammed G-K, Hassan H. Validity and reliability of the Arabic version of the Female Genital Self-Image Scale. **J Sex Med 2014;11:1193-1200.**
57. Nimbi F, Tripodi F, Simonelli C, et al. Sexual Modes Questionnaire (SMQ): Translation and psychometric properties of the Italian version of the Automatic Thought Scale. **J Sex Med 2018;15:410-415.**
58. Nowosielski K, Wróbel B, Sioma-Markowska U, et al. Development and validation of the Polish version of the Female Sexual Function Index in the Polish population of females. **J Sex Med 2013;10:386-395.**
59. Opperman EA, Benson LE, Milhausen RR. Confirmatory factor analysis of the Female Sexual Function Index. **J Sex Res 2013;50:29-36.**
60. Pakpour A, Zeidi I, Saffari M, et al. Psychometric properties of the Iranian version of the Sexual Quality of Life Scale among women. **J Sex Med 2013;10:981-989.**
61. Pakpour A, Zeidi I, Ziaeiha M, et al. Cross-cultural adaptation of the Female Genital Self-Image Scale (FGSIS) in Iranian female college students. **J Sex Res 2014;51:646-653.**
62. Pascoal P, Alvarez M-J, Pereira C, et al. Development and Initial Validation of the Beliefs About Sexual Functioning Scale: A gender invariant measure. **J Sex Med 2017;14:613-623.**
63. Rehman K, Asif Mahmood M, Sheikh S, et al. The Female Sexual Function Index (FSFI): Translation, validation, and cross-cultural adaptation of an Urdu version "FSFI-U." **Sex Med 2015;3:244-250.**
64. Rellini A, Meston C. The sensitivity of event logs, self-administered questionnaires and photoplethysmography to detect treatment-induced changes in female sexual arousal disorder (FSAD) diagnosis. **J Sex Med 2006;3:283-291.**
65. Rillon-Tabil N, Malong C, Vicera J, et al. Translation and validity of the female sexual function index Filipino version (FSFI-Fil). **Philippine J Intern Med 2013;51.**
66. Rogers R, Rockwood T, Constantine M, et al. A new measure of sexual function in women with pelvic floor disorders (PFD): The Pelvic Organ Prolapse/Incontinence Sexual Questionnaire, IUGA-Revised (PISQ-IR). **Int Urogynecol J 2013;24:1091-1103.**
67. Rosen R, Bachmann G, Reese J, et al. Female Sexual Well-Being Scale (FSWB Scale): Development and psychometric validation in sexually functional women. **J Sex Med 2009;6:1297-1305.**
68. Ryding E, Blom C. Validation of the Swedish version of the Female Sexual Function Index (FSFI) in Women with Hypoactive Sexual Desire Disorder. **J Sex Med 2015;12:341-349.**
69. Selcuk S, Kucukbas M, Cam C, et al. Validation of the Turkish version of the Sexual Health Outcomes in Women Questionnaire (SHOW-Q) in Turkish-speaking women. **Sex Med 2016;4:e89-e94.**
70. Sidi H, Abdullah N, Puteh S, et al. The Female Sexual Function Index (FSFI): Validation of the Malay version. **J Sex Med 2007;4:1642-1654.**
71. Sills T, Wunderlich G, Pyke R, et al. The Sexual Interest and Desire Inventory-Female (SIDI-F): Item response analyses of data from women diagnosed with hypoactive sexual desire disorder. **J Sex Med 2005;2:801-818.**
72. Stephenson K, Toorabally N, Lyons L, et al. Further validation of the Female Sexual Function Index: Specificity and associations with clinical interview data. **J Sex Marital Ther 2016;42:448-461.**
73. Sun X, Li C, Jin L, et al. Development and validation of Chinese version of Female Sexual Function Index in a Chinese population—A pilot study. **J Sex Med 2011;8:1101-1111.**
74. Takahashi M, Inokuchi T, Watanabe C, et al. The Female Sexual Function Index (FSFI): Development of a Japanese version. **J Sex Med 2011;8:2246-2254.**
75. Ter Kuile M, Brauer M, Laan E. The Female Sexual Function Index (FSFI) and the Female Sexual Distress Scale (FSDS): Psychometric properties within a Dutch population. **J Sex Marital Ther 2006;32:289-304.**
76. Trudel G, Dargis L, Cadieux J, et al. Validation of the Female Sexual Function Index (FSFI) and presentation of norms in older women. **Sexologies 2012;21:161-167.**
77. Trutnovsky G, Nagele E, Ulrich D, et al. German translation and validation of the Pelvic Organ Prolapse/Incontinence Sexual Questionnaire—IUGA revised (PISQ-IR). **Int Urogynecol J 2016;27:1235-1244.**
78. Vallejo-Medina P, Perez-Duran C, Saavedra-Roa A. Translation, adaptation, and preliminary validation of the Female Sexual Function Index into Spanish (Colombia). **Arch Sex Behav 2018;47:797-810.**
79. Velten J, Scholten S, Graham CA, et al. Psychometric properties of the Sexual Excitation/Sexual Inhibition Inventory for Women in a German sample. **Arch Sex Behav 2016;45:303-314.**
80. Verit F, Verit A. Validation of the Female Sexual Function Index in women with chronic pelvic pain. **J Sex Med 2007;4:1635-1641.**
81. Wang H, Lau H-H, Hung M-J, et al. Validation of a Mandarin Chinese version of the pelvic organ prolapse/urinary incontinence sexual questionnaire IUGA—revised (PISQ-IR). **Int Urogynecol J 2015;26:1695-1700.**
82. Wiegel M, Meston C, Rosen R. The Female Sexual Function Index (FSFI): Cross-validation and development of clinical cutoff scores. **J Sex Marital Ther 2005;31:1-20.**
83. Witting K, Santtila P, Jern P, et al. Evaluation of the Female Sexual Function Index in a population based sample from Finland. **Arch Sex Behav 2008;37:912-924.**

84. Wolpe R, Queiroz A, Zomkowski K, et al. Psychometric properties of the Female Sexual Function Index in the visual analogue scale format. **Sex Health 2017;14:213-220.**
85. Wylomanski S, Bouquin R, Philippe H-J, et al. Psychometric properties of the French Female Sexual Function Index (FSFI). **Qual Life Res 2014;23:2079-2087.**
86. Zachariou A, Filiponi M, Kirana P. Translation and validation of the Greek version of the Female Sexual Function Index questionnaire. **Int J Impot Res 2017;29:171-174.**
87. Zohre M, Minoo P, Ali M. Pelvic Organ Prolapse/Urinary Incontinence Sexual Questionnaire (PISQ-12): psychometric validation of the Iranian version. **Int Urogynecol J 2014; 26:433-439.**
88. Chedraui P, Pérez-López FR, Sánchez H, et al. Assessment of sexual function of mid-aged Ecuadorian women with the 6-item Female Sexual Function Index. **Maturitas 2012; 71:407-412.**
89. Lee Y, Lim MC, Joo J, et al. Development and validation of the Korean version of the Female Sexual Function Index-6 (FSFI-6K). **Yonsei Med J 2014;55:1442-1446.**
90. Mitchell K, Ploubidis G, Datta J, et al. The Natsal-SF: A validated measure of sexual function for use in community surveys. **Eur J Epidemiol 2012;27:409-418.**
91. Pérez-López FR, Fernández-Alonso AM, Trabalón-Pastor M, et al. Assessment of sexual function and related factors in mid-aged sexually active Spanish women with the six-item Female Sex Function Index. **Menopause 2012;19:1224-1230.**
92. Bartula I, Sherman KA. Development and validation of the Female Sexual Function Index adaptation for breast cancer patients (FSFI-BC). **Breast Cancer Res Treat 2015; 152:477-488.**
93. Burri A, Cherkas L, Spector T. Replication of psychometric properties of the FSFI and validation of a modified version (FSFI-LL) assessing lifelong sexual function in an unselected sample of females. **J Sex Med 2010;7:3929-3939.**
94. Carvalho J, Vieira AL, Nobre P. Latent structures of female sexual functioning. **Arch Sex Behav 2012;41:907-917.**
95. Bland JM, Altman DG. Statistical methods for assessing agreement between two methods of clinical measurement. **Lancet 1986;327:307-310.**
96. Bland JM, Altman DG. Agreement between methods of measurement with multiple observations per individual. **J Biopharm Stat 2007;17:571-582.**
97. Carvalheira AA, Brotto LA, Leal I. Women's motivations for sex: Exploring the diagnostic and statistical manual, fourth edition, text revision criteria for hypoactive sexual desire and female sexual arousal disorders. **J Sex Med 2010; 7:1454-1463.**
98. Balon R, Segraves RT, Clayton A. Issues for DSM-V: Sexual dysfunction, disorder, or variation along normal distribution: Toward rethinking DSM criteria of sexual dysfunctions. **Am J Psychiatr 2007;164:198-200.**
99. Gierhart BS. When does a "less than perfect" sex life become female sexual dysfunction? **Obstet Gynecol 2006; 107:750-751.**
100. Sungur MZ, Gündüz A. A comparison of DSM-IV-TR and DSM-5 definitions for sexual dysfunctions: Critiques and challenges. **J Sex Med 2014;11:364-373.**
101. de Vet HC, Terwee CB, Ostelo RW, et al. Minimal changes in health status questionnaires: distinction between minimally detectable change and minimally important change. **Health Qual Life Outcomes 2006;4:54.**
102. Puppo V, Puppo G. RE: Bartula I, Sherman KA. Development and validation of the Female Sexual Function Index adaptation for breast cancer patients (FSFI-BC). **Breast Cancer Res Treat 2015;153:705-706.**

SUPPLEMENTARY DATA

Supplementary data related to this article can be found at 10.1016/j.jsxm.2019.03.001